DOCUMENT RESUME

ED 105 031 UD 015 021

AUTHOR	Cahill, Robert J.; Foley, Joseph J.
TITLE	Evaluation and Evaluative Research in an Urban Bilingual Program.
INSTITUTION	Heuristics, Inc., Dedham, Mass.
PUB DATE	73
NOTE	23p.; Not available in hard copy due to marginal legibility of original document
EDRS PRICE	MF-$0.76 HC Not Available from EDRS..PLUS POSTAGE
DESCRIPTORS	Academic Achievement; Bilingual Education; Case Studies; Educational Research; *Evaluation Methods; Parent Participation; *Program Evaluation; *Research Problems; Spanish Speaking; Student Adjustment; Teacher Behavior; *Urban Education; Urban Youth
IDENTIFIERS	*Elementary Secondary Education Act Title VI; ESEA Title VI

ABSTRACT

This document discusses an evaluation of the second year of a bilingual program in a large Eastern city funded by Title VII of the Elementary and Secondary Education Act. Like those in a number of other localities, the Spanish-English program described here developed from earlier "English as a Second Language" project. The evaluation was highly localized in nature, in contrast to the contracting process at the national level. Other circumstances served to emphasize the program's local aspect. There was first the matter of a distinctive pattern and rate of influx of Spanish-speaking students. Secondly, the total staff experience in the years preceding necessarily assumes unique configurations in each urban setting. A third circumstance involves the recent history of evaluation in a given setting. Because the present contractor had been hired part way through the first year of the program, the evaluation report for that first academic year was protected if not hedged--by statements of the constraints imposed by the timing of his appearance on the scene. The second year evaluation sought to determine the extent of student improvement in both the cognitive and effective domains. Then the evaluation went further, looking at parental involvement, teacher performance, and case studies related to those students who participated in a special Social Readjustment Class (Author/JM)

ED105031

EVALUATION AND EVALUATIVE RESEARCH IN AN URBAN BILINGUAL PROGRAM

Robert J. Cahill
Joseph J. Foley*

Heuristics, Inc.

Heuristics, Inc. has specialized in independent ("third party") evaluations in local school systems. The firm has conducted over 250 evaluations of programs funded under Titles I, II, VI, and VII of the Elementary and Secondary Education Act. The officers and other professionals in the company have conducted various research projects in the field of evaluation; continuing corporate interest in evaluative research is reflected in the present chapter.

---

* Dr. Cahill supervised the evaluation described in the chapter; Dr. Foley is a member of the board of directors, and special consusltant on computational linguistics. Dr. Cahill is Vice President of Heuristics, Inc., and Dr. Foley is Associate Professor, Boston State College.

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

UD 015021

INTRODUCTION

Bilingual education has become an important component of the "compensatory" education activities sponsored either directly or indirectly by the federal government. Although bilingual education covers many groups for whom English is, in effect, a second language, its primary focus has been on the Spanish speaking population in the Southwest, and in the large urban centers. This chapter reports on a local program covering a number of elementary schools in a large Eastern city. The evaluation sought to determine the extent of student improvement in both the cognitive and affective domains. This is of course the traditional approach, but the evaluation went further, looking at parental involvement, teacher performance, and case studies related to those students who participated in a special Social Readjustment Class.

PART I: CONTEXT OF THE EVALUATION

A. A LOCAL RATHER THAN A NATIONAL EVALUATION

This chapter has as its subject an evaluation of the second year of a bilingual program in a large Eastern city, funded by Title VII of the Elementary and Secondary Education Act. Like those in a number of other localities, the Spanish-English program described here developed from earlier "English as a Second Language" (ESL) projects--projects extending back to the middle '60's. Following publication, in 1968, of guidelines for Title VII proposals, the administrators who had organized the ESL projects in the school system prepared formal applications, and received a grant for the city's first bilingual program. The contractor had completed work for the earlier ESL program, and was retained initially as project auditor for the new bilingual project.

The evaluation was highly localized in nature; the contract was sought from, and approved by, local school officials, in contrast to the contracting process at the national level. A particular program in a single metropolitan school system was involved, such a circumstance concentrates the evaluator's attention.

Other circumstances served to emphasize the program's local aspect. There was first the matter of a distinctive pattern and rate of influx of Spanish-speaking students--a pattern and rate not likely to be closely similar to others across urban school systems, since demographic variables are affected by geographic differences. Secondly, the total staff experience in the years preceding necessarily assumes unique configurations in each urban setting. Varying policies on staff rotation, on consultants, and on hiring and dismissal will give each city some measure of individuality for such emerging national projects as those in bilingual education. A third circumstance accenting the "local" nature of the program involves the recent history of evaluation in a given setting. Such a variable as the identity of the most recent evaluators (or the very novelty of organized assessment in a system previously without such services) will influence an evaluation differentially across locales. The point might be more specifically made by noting that the present contractor, having begun work as auditor, assumed the role of evaluator mid-way through the first year of the program--the evaluation to that point having been the responsibility of a university-based consultant. The reason for the change involved demands on the previous evaluator's time which increased beyond his availability. Just such details of recent

history comprise the variable described in the third "localizing" influence. Differences, then, involving the subjects, the staff, and the evaluators combined to contribute a highly localized character to the program.

That the evaluation demands on the university consultant grew beyond what he could satisfy does not necessarily reflect discredit on the design or scheduling of his assessment. The change in evaluator is explained rather by the need in such a program for a degree of flexibility in time not easily compatible with the university commitments of an evaluator-consultant. There has clearly been an effective role in program evaluation for university faculty, and such a role will and should continue. The more complex the program, however, the greater the risk that the evaluation--especially the formative evaluation--may present unpredicted problems (and opportunities) that must be met when they arise, and not at previously scheduled times. The unpredictability applies as well to the consultant's university responsibilities. His service to a program is usually arranged for in advance; he is only too likely to find that academic demands increase in unplanned ways--doctoral "advisees" are added, new committees are formed, and so forth. September's faculty responsibilities can and often do undermine projections made during the preceding spring.

B. PREVIOUS ASSESSMENT OF THE PROGRAM

Because the present contractor had been hired part way through the first year of the program, the evaluation report for that first academic year was protected--if not hedged--by statements of the constraints imposed by the timing of his appearance on the scene. ("Accordingly, some areas have received much attention, and others very little, as dictated by the priorities assigned.") The acceptable compromise in such situations is one which puts a premium on attention to student outcomes. While properly analyzed evaluative data on staff, administrative, and community variables will usually return additionally useful information to a program, it is clear that when desired student outcomes appear, the interaction among the other variables may be presumed to have been essentially effective. The main burden of the report for the preceding year was an accounting of what changes were effected in the participants; and the data on such matters as program management and parent-program interaction were gathered and analyzed more for descriptive than prescriptive purposes.

Because of the late award of the contract for the 1969-70 evaluation (that is, the assessment for the year preceding the

work discussed in these pages), pre-testing took place in January and February of 1970: the time available for demonstrating gains was therefore from four to five months. Despite the limited period available for measurement of cognitive growth, the data revealed significant increases in two of three language abilities specified in the objectives. Computation skills were not pre-tested; end-of-program measures, however, were compared with national norms. Data on affective outcomes were gathered through anecdotal records, and were analyzed in light of such relevant information as selected home background variables and the rate of absenteeism. This characteristically more difficult measurement domain held even less prospect for discovery of observable gains in the time available, given the slow and elusive nature of attitudinal change.

Since choice of instruments constituted an importnat aspect of the evaluation which this chapter covers, there is reason to consider briefly one aspect of that same matter in the previous (1969-1970) evaluation.

The report for that year mentions at one point the evaluators' effort to develop and employ a measure of language fluency. The technique attempted was chosen by the contractor as one expecially appropriate for a program which sought improvement of oral-aural abilities: it involved the use of the type-token ratio (TTR). The ratio of different words (types) to total words (tokens) is a fluency measure, and the relevant counts are amenable to statistical analyses.[1] To generate the spoken protocols, the examiners developed a standardized set of photographs, designed to elicit segments of speech by picturing people, places, and things familiar to all the participants. In pilot testing, the graphics did not succeed in eliciting long enough units of discourse for TTR analyses. Nevertheless, at the time of this writing, the evaluators still feel that the choice of instrument was especially appropriate, and further believe that had there been resources for further development and pilot testing, the potential in the techinque might have been realized. (An implication for cultural study is described in Part Two of the present chapter.)

The purpose in so describing one comparatively small aspect of the instrumentation is to document one comparatively large aspect of the evaluator's role--his responsibility, namely, to maximize the learning which an effective test can promote. Staff understanding of what speech-fluency quantitatively implies may in fact be enhanced by TTR testing. When the ratio is separately computed for successive segments of n words (the decremental TTR's), the instructor has a measure of the rate at which the

learner "burns up" his available vocabulary. It is clear that such analyses are not common in language teaching--a circumstance not to be wondered at, in light of the tedious nature of TTR computations done by hand. The computer, however, calculates and reports the ratios with instant speed, and reports as well on the range and nature of vocabulary, since the data can be machine alphabetized. There follows from all this the important additional benefit of transferring insights from the testing techniques to the instructional practice. To the extent that an evaluation includes that kind of effectiveness, it has overcome the stultifying influence of the simple prettest-posttest model for assessment.

C. CONTRACTOR-CLIENT RELATIONS

The working relationship between the contractor and client was influenced in the present instance by the fact that earlier contracts had been awarded and carried out for study by ESL programs conducted by the same client. Although the bilingual program involved a number of school administrators who had not worked in the earlier ESL projects, some members of the ESL staff had continued in similar positions when the new program began. The sense of shared experience was not least among the factors contributing to the needed rapport.

Important also to the contractor-client relationship was the range of bilingual ability among the evaluation staff. It is seldom possible to put into the field as many bilingual evaluators as are needed--nor was the present study an exception. Among the evaluators who spoke only English, there were occasions when the interaction with the teachers disappointed expectations on both sides. One major implication for future planning follows from such experience: it is difficult to emphasize enough the need for informing (if not convincing) the project staff that the evaluation is as challenging as it is centrally important. "The great need is for the most advanced thinking about evaluation to be communicated to the bilingual staff so that the statement of aims, the teaching, and the testing may all be correlated."[2]

D. POINTS OF UNIQUENESS ABOUT THE EVALUATION

Two aspects of the study--one involving parents and the other the teachers--deserve mention as contributing some measure of uniqueness.

Of special interest, first, are not only the findings on parental involvement, but also the process of establishing a particular level of involvement as satisfactory. There had been no reliable guideline prepared for predicting or prescribing levels of parental involvement. Title VII had funded comparatively few projects, and the distinctive pattern of home-and-school relationships found in bilingual programs had never adequately been profiled. Unlike the staff earlier associated with ESL projects, the bilingual teachers accepted responsibility for discussing with the parents, in the parent's native language, a full variety of school subjects. The expectations concerning interaction with the parents were to some extent being investigated, therefore, at the same time that their fulfillment was being measured.[3]

A second point of uniqueness concerned the fact that the design of the evaluation included measures on one variable not often addressed specifically in such studies--that of teacher performance. It may be that certain theoretical difficulties inhibit work in this direction; if so, chief among the theoretical issues is that of the discriminability of teaching and learning. The kind and scope of measures on teacher performance will depend, to a considerable extent, on what position is taken on certain key questions, of which the following is representative: Does the definition of teaching include activities not causally related to learning? One answer will encourage the gathering of data of a kind that other answers would hold irrelevant. The strategy governing the present design made the point moot, by requiring that the data on the teachers include information on a wide variety of activities, without worrying the question of what causal relationship each kind of data had to the learning that took place. That measures of teacher performance were taken gave a comparatively unusual cast to the assessment.[4]

E. NEEDED IMPROVEMENT IN EVALUATION DESIGN

Among the controlling assumptions incorporated in Title VII are those which emphasize that bilingual programs permit clearer distinctions to be made between what is taught, and language as medium for that content. The focus can more clearly be placed on the learning, since the truism may still be trusted that learning succeeds best when lessons are conducted in one's mother tongue. In theory, therefore, the program might have been expected to foster such desirable affective outcomes as improvement in attitude toward school, toward self, and toward self-as-learner. The assessment was intended to document the amount and kind of attitudinal change made possible

by insulating the learners from the frustration of filtering their school efforts through the tangled screen of English. As detailed later, only part of this "evaluative research" function was fulfilled. The explanation closest to hand does not adequately account for the discrepancy between intended and achieved results: the problem, namely, of the dearth of effective attitudinal instruments, which is as real as it is persistent. To that deficiency must be added a design flaw, which permitted loss of opportunities to orchestrate the gathering of affective data across the program components. Home visits, for example, were made regularly; and a modest amount of attitudinal information was gathered from the perceptions of the staff who made the visits. Collecting affective data in this unstructured fashion, though productive to some extent, could (and did) prevent realizing the full potential of available data. Given the already well recognized problems with affective measures, effort might well have been made to integrate more closely the instruments and procedures which <u>were</u> available.

PART II: THE EVALUATION

A. GENERAL EVALUATIVE TECHNIQUE

Program operations took place in seven classes and in three schools. The highly transient nature of the student population restricted usable pre-post data to an n of about 70. The cognitive goals included the following areas: speaking and listening skills in English and Spanish, reading in English and Spanish, mathematics, and science. Affective goals included the areas of attitude toward school, toward self-as-learner, and self concept. The reader should note that a Social Readjustment Class was a subcomponent of the bilingual program.

The overall evaluation strategy was to combine the advantage of statistical analysis of results of standardized instruments, with less formal (though still structured) techniques using observers. Case studies were completed for a small sample of children (5) in one component--the Social Readjustment Class. Because the details of the case studies could be incorporated in the report, there was the additional advantage of having evaluative data preserved in a form that would allow for further analysis by interested readers of the final report.

B. OBSERVATIONAL DATA

Site selection for the assessment posed no problem, given the few schools involved. This circumstance influenced the decision to train observers, since there was reason to believe that the data they gathered would amply repay the training and observation time involved. In retrospect, it appears that the average of nine visits to each class was only marginally sufficient to give the observers the degree of "invisibility" needed to prevent their presence from influencing what they were there to observe. The efficiency curve turns dramatically upward when the unobtrusive point is reached: the nine visits did not allow this exponential rate of data-return to take full effect. During their deliberations on the relation between frequency of observation and fund of data returned, the evaluators were mindful of two extremes, the first represented by the year-long observation-studies by Jackson reported in his Life in Classrooms, and the second represented by the discussion of minimal requirements in Webb's Unobtrusive Measures.[5]

As an observer Jackson succeeded in achieving almost total "invisibility" by means of long-term attendance in the classrooms under study. Because his presence came to be taken for granted,

he very seldom by his act of observing changed the nature of the classes he observed. This same interference phenomenon (which is a translation of the physicists' "uncertainty principle" to social science) explains the wording of the second of the titles mentioned: Unobtrusive Measures. In that volume Webb and his co-authors surveyed widely-varied sources of "non-reactive" data--information which can be gathered in such a way that its collection does not modify the conditions being studied. Their work helped define by example some threshold points beyond which the presence and activity of an observer no longer obtrude.

C. INSTRUMENT SELECTION/IMPLICATIONS FOR RESEARCH

In the selection and development of instruments, the predictable considerations were allowed to control the decisions: program goals, characteristics of the participants, and technical characteristics of the instruments available (reliability, validity, norm data, etc.). For one part of the cognitive testing, the identity of the norm groups and the recent date of the normative studies were reflected in the choice of the Metropolitan Achievement Tests (1970). A very recent survey of the country's 100 largest school districts finds that battery first among the top five mental ability and achievement tests.[6] Selection of The Test of Aural Comprehension (English Language Institute, University of Michigan) was influenced by the extensive theoretical support published by Robert Lado and other researchers associated with that instrument. The latter point deserves some elaboration.

Although the theory does not purport to be inclusively applicable to instruction involving two languages, the "contrastive analysis" approach has held the field now for more than a decade; it dates from Robert Lado's publication Linguistics Across Cultures.[7] Briefly stated, contrastive analysis suggests that schooling which involves a native and a target language will profit most when primary attention is given the points at which the two languages contrast. The theory proposes that the degree to which the linguistic structures differ defines the degree of difficulty to be expected in learning the target language. It is precisely this predictive value that most strongly recommends the theory (and tests which reflect the theory) to bilingual programs. Contrastive analysis holds much promise for the training of new teachers for bilingual programs. Older teachers in the field are by long experience sensitized to the points of "interference" in the two languages. Because bilingual schooling was described as late as 1970 as being "still in its infancy,"[8]

any theoretical work which helps new teachers anticipate the major difficulties in bilingual programs is considerably important on that ground alone.

These remarks on the theoretical background which influenced a choice of test provide a further example of the ways in which the evaluator's activities can serve a research purpose. Because the program provided opportunity for the staff to consider the implications of the emphases in the tests, they had an opportunity to include the results of contrastive analysis in instruction. Easily the most useful guidance they thus received was the proportionate time to be given various kinds of language drills.

An evaluation inevitably precipitates "fall-out" not only of this instructional kind, but also of a kind affecting program administration and community relations. It will be seen that an assessment has "in-progress" influence beyond the benefits intended by the details of interim reports in the formative evaluation phase. (The now-familiar distinction between "formative" and "summative" evaluation bears reemphasis: the former term includes those activities whose results are returned to the program staff in time to have an effect on end-of-program outcomes; summative evaluation refers to activity conducted after a program has ended.)

Among the technical characteristics of instruments selected were reliability and validity indices. The criteria of reliability and validity comprise the most frequently studied technical aspects of test usage. A test is reliable when it consistently measures what it is supposed to measure; it is valid when it accurately measures what it is supposed to measure. (Each term admits of various classifications: reliability is severally defined by the statistical operations used in its determination--e.g., "split-half," "test-retest," etc.; validity is similarly classified by such methods of determining it as "concurrent," "face," "content," etc.)

The manuals for the Metropolitan and The Test of Aural Comprehension report reliability coefficients, and validation studies. Because the evaluators had conducted the assessment for the preceding (first) year of the program, there had been opportunity to conduct limited study of test reliability for the certain other instruments which continued to be appropriate for use in the second year, given the similarity of program goals in the two years. For the Comprehension of Oral Language (Inter-American Series, Guidance Testing Associates), the reliability coefficient thus computed with data from the first year was .96.

Correlation coefficients between two forms of the Spanish edition of Prueba de Lectura were also computed during the first year of the program; they were .86 for Level 1, .74 for Level 2, and .90 for Level 3.

It will be clear from the foregoing that a variety of commercially available measures were employed: a balanced evaluation design will often provide as well for the development of at least one measure by the program staff. The program administration took advantage of the opportunity to include staff training in evaluation among the benefits to be derived. The science goals served as the immediate context in which this training was accomplished. With some assistance from the evaluator, the Science Curriculum Coordinator prepared a measure with a satisfactory degree of content validity--defined as the accuracy and representativeness with which a test samples the subject matter and behaviors included in the instruction.

As mentioned earlier, the bilingual approach suggests that some degree of attitudinal gain should attach to the use of the student's native language for instruction. Accordingly, one of the two attitudinal measures included both teacher and student reactions on the matter of the students' attitude toward themselves as learners. The cognate issue of change in self concept was tested through the use of the Piers-Harris Self-Concept Scale.

Although the synopsis of outcomes has not yet been given, it seems appropriate to comment here on a design improvement which would be carefully considered if the evaluation were to be done again. The point has to do with rather a simple expedient--namely that of alerting the classroom observers to the actual items on the affective measures used. (Other staff members need the same information, too; but the classroom observers need it most.) Although the observation checklists were comparatively "busy" ones, it appears with the usual clarity of hindsight that the observers might have been able to attend as well to some of the precise attitudinal signals included in the affective instruments. As important as any other benefit from such a design feature would be the opportunity to integrate the findings in the several sections of the resulting report.

The evaluators feel that another lesson deserves to be overlearned, concerning the final report of an evaluation project. This is, that a report should be a highly interactive product, and not a collection of sub-sections which seem unaware of each others' existence. In earlier studies the evaluators attempted to satisfy this standard by means of recommendations and conclusions which in one final section touched on the sections that came before; the result must yield to that kind of final report which shows that it knows on one page what another page will say,

as well as what another page has said. It has been noticed that a disproportionately large dividend in reader-reassurance follows from the simple techniques of including occasional anticipatory remarks in the pages of what is so often the single tangible outcome of the evaluation.

D. SYNOPSES OF FINDINGS

The major cognitive findings are synopsized in Table 1, which follows. A number of the implications have been touched on previously.

Table 1

Cognitive Outcomes-Summary

Area Measured	Significant Gain
Oral English Production	x
Aural English Comprehension	x
Reading--English	x
Reading--Spanish	
Vocabulary	x
Comprehension	x
Speech	x
Aural Spanish	x
Mathematics	X
Science	x

Affective outcomes showed less impressive gains, as documented by the results in Table 2. Only in the matter of attitude toward self-as-learner did the changes reach a level of statistical significance, as determined by t tests of correlated data (i.e., matched pairs of pre-post scores).

The significant differences reported in the tables are encouraging; there need be no danger, however, of over-reading

the statistical results. The several other kinds and sources of data combine to support the statistical judgment that the program achieved satisfactory levels of overall effectiveness.

Table 2

Affective Outcomes - Summary

Area Measured	Significant Gain	Not Significant
Attitude Toward School (Teacher Rating)		x
Attitude Toward School (Student Response)		x
Attitude Toward Self-As-Learner (Teacher Rating)	x	
Attitude Toward Self-As-Learner (Student Response)		x
Self-Concept		x

E. HYPOTHESIS-GENERATING FUNCTION FOR DATA REPORTED

Findings of a less generalizable kind followed from the use of the case study approach for the program component known as the Social Readjustment Class. Considered in the context of all the data gathering techniques employed, the case study affords an especially clear illustration of the "trade-off" that follows from a choice of evaluative method. The technique involved a willing forfeit of generalizability of the results, in favor of the in-depth nature of what is found thereby. As mentioned earlier the several case studies were reproduced in the report, for the purpose of providing not only some specific data, but some material for further study. Here again the evaluation design accepts the opportunity to serve a research purpose, since the content of the case studies may generate new hypotheses important to the program. The narrative portion of one of the cases is reprinted here, with the foregoing observations in mind:

Behavior-previous school year.

E____ had two years of kindergarten as part of a readiness program to better prepare children for school. He was first enrolled in September, 1968. One of E____'s major problems was truancy. In KI he attended school 29 days out of 180. In KII he attended school 29 days out of 180. When present he was usually very quiet and withdrawn. On occasion he would explode: climbing over the tables, throwing blocks, swinging at others and screaming. He would sometimes play with a puzzle while other children did paper work. He never participated in group activities or communicated with the teacher or other students. He would "freeze" and act frightened when the teacher approached him. The teacher cannot recall him ever smiling.

Behavior-prior to referral.

E____'s behavior prior to referral was described as frightened and antagonistic. He would wander around the room touching things and messing up the other students' work. He had a great deal of difficulty getting along with others. He did not seem to understand that anything that he wanted was not his to have. He could not participate in a group activity and never completed an assignment or work sheet. He could not follow directions for behavior control or work. When the teacher attempted to gain his cooperation in following routines of the school day he would physically fight with her or run. As in kindergarten, truancy was a problem. He was absent seven days out of sixteen during the month of September. It was impossible to make an academic assessment at this time.

Procedure following referral.

E____ was referred at the beginning of October. The project teacher's initial goal was to improve E____'s rate of school attendance. She checked each morning with E____'s teacher and if he was not there she went to his home. The teacher explained how important school was for E____ and told the mother that she would come and check when E____ was absent. On days when she was busy with another student in the morning she visited E____'s home

in the afternoon. E_____ had seven absences in October. The teacher brought him to school four times. In December E_____ was absent one time for illness and the project teacher had to bring him once. For the first three months of 1971 he was absent eight times, all of them because of illness. He attended school voluntarily on the other days.

In the beginning he had to be forcibly removed from the regular classroom by the project teacher when trouble occured. He usually remained in his regular classroom only for opening exercises, lunch time, and a short time prior to dismissal. The project teacher attempted to teach him how to behave in class by describing the activities and explaining what he could expect to do when he returned to class. She described to him exactly what he should do when returning and practiced these procedures in the bilingual classroom. E_____ then began telling her what he would do when he returned to his classroom. He began to understand organization and routines and their application to daily activities.

The project teacher discovered that one of E_____'s problems was setting extremely high unrealistic goals for himself. He seemed to think that any effort would be downgraded and ridiculed. When working on paper he would scribble all over it even when done correctly. When working on the board he would continually erase and was never satisfied with what he did. The project teacher continually reassured him that any effort he made at school work would be acceptable and praiseworthy. About two days after one such encouragement he admitted liking a picture which he had drawn. The next morning he asked to remain in the classroom. An agreement was made that he could stay in the room if he obeyed the teacher. After one failure, he managed to control his behavior.

Because he was so far behind academically, the project teacher started spending one-half hour per day with him in the regular room to help him learn how to do worksheets and catch up work he had missed. As soon as he became familiar with what procedure was used on each worksheet he worked independently.

Behavior-present

E____ still occasionally has trouble with directions given to the whole group. At these times the project teacher spends some time with him explaining any new procedures.

He only has to be removed from his room about once every three weeks for an hour because of disruptive behavior. His regular teacher and speech teacher have noticed that E____ is much more relaxed and communicative. His mother reported that his behavior has improved at home.

The project teacher no longer needs to go to his home and bring him to school.

Academically he is at the lower level of the group. He can add single numbers with difficulty and is reading at the pre-primer level.. He does not always actively participate in such group activities as music or storytelling, but he does join in the group without disrupting it.

Evaluation

Prior to inclusion in this project, E____'s truancy, disruptive behavior, and inability to communicate or follow directions had prevented him from deriving any benefits in the school situation. Constructive, remedial help was given to him by this project and he is now a participating member in his regular classroom.

It should be emphasized that the case study approach had as its first purpose the gathering of evaluative data on the outcomes of the component mentioned above--the Social Readjustment Class. In that sense it serves a hypothesis-function--the hypothesis here might be stated in null-form: in the regular classroom there is no difference in behavior between students referred to the Social Readjustment Class, and others not so referred. The point at issue here, however, is the hypothesis-generating function that is filled by the case studies, and by including them in the final report.

To illustrate this function one example may be developed. The close reader of the narrative might be drawn to the several ways in which the "routine" aspects of the classroom became useful

to the teacher in assisting this student. The "routines of the school day" are referred to under the second sub-heading; the student's own routine is implied in the second paragraph under the third sub-heading "he usually remained..only for opening exercises...(emphasis supplied here, and in other quotes); the teacher practiced these "procedures"; "routines" are again mentioned under the third sub-heading, and "daily activities"; when "he became familiar with what procedure was used on each worksheet he worked independently." The overall impression gathers quickly that the student described was one who stood a very good chance to profit considerably by learning to rely on the stability of the classroom environment. The hypothesis which follows, then, would be one which proposed: When instructional activities and curricular materials specifically encourage the referred student to perceive the classroom as a stable environment, disruptive behavior will be diminished.

F. CONCERN FOR ACTIVITIES OTHER THAN GOAL-DIRECTED

It is unlikely that the data can be relied upon to support other kinds of generalizations. But it is important to recognize that the study did accept the wider responsibilities that every comprehensive design inevitably involves. In an article titled "The Politicization of Evaluation," Carol H. Weiss observes that "the system model (of evaluation) recognizes that organizations engage in activities other than the achievement of their goals. A study, therefore, should not focus exclusively on goal attainment but should look on measures on the effectiveness of other organizational functions, such as recruiting resources, maintaining the structure, achieving integration into the environment."[9] With these considerations in mind, the report here under study attends, for example, to recruitment activities--such as the highly successful recruiting done by the Coordinator of the Parent Tutorial Program. This wider angle of focus goes beyond the now-familiar concern for unintended and undesired program outcomes, to the system concern of transactional evaluation, described by Rippey.[10]

G. IMPLICATIONS FOR RENEWED EVALUATION OF SIMILAR PROGRAMS

As important as what the findings say, is what they do not say. The single most serious silence in the report is the absence of any significant work with spoken data. Bilingual programs are designed to conserve valuable cultural traits and differences, and they raise very conscious and specific safeguards against the kind of "Americanization" that is destructive of the values of other cultures. Many such values are incorporated in language

habits--and the disproportion in output between speech and writing need not be argued. It follows that some of the best data for accenting differences in values across cultures is available in the student's oral language. These considerations were among those which led to the work with TTR's (type-token ratios) described earlier as having been done in the previous year's evaluation.

What it comes down to is one more problem with instrumentation. The samples of oral discourse must be collected with care, if the analysis is to include concern for cultural signals encoded in speech. The more natural the setting and the experience, the more reliable are the interpretations of the cultural values revealed by the language habits. "The linguists' requests to a speaker to produce particular forms of languages may be interpreted by speakers as a request to behave in a certain way (i.e., to be formal or informal), or to emphasize otherwise some aspect of his social personality--The [contrived] setting...tends to reduce to a minimum the very type of linguistic devices by which social meaning is conveyed..."[11]

What is essential, then, is to gather the oral data in as natural a setting as possible. Initial effort with TTR's involved a single graphic stimulus, as noted in part one of this chapter. Pilot testing had not sufficiently cleared up the problems to warrant use of the technique in the present study. If the evaluation were to be repeated, the "naturalness" of the setting in which speech was sampled would be a major concern. (One expedient: the students might tape some songs, and the taping would ostensibly be for the purpose of preserving the songs; what in reality would be wanted would be the conversations and other segments of speech occurring between the songs.) Operationalizing the methodology, clearly, is a good deal more difficult to manage than a description of desired results.

Two hopeful developments combine to encourage gathering this kind of evaluation data. For the first, audio cassette technology has continued to show refinements and greater accessibility; cassette duplicating equipment is also now more easily and more economically obtainable. The second technological advance involves time-impressed speech. Samples of oral data can be re-recorded in such fashion that listening time can be reduced by fifty-percent, without distortion. (The original tapes are not lost in the process.) When large samples of speech are drawn, some such economy is essential, if only in helping reduce the time needed to transcribe the spoken data. It should be noted however that time-compressed speech would not be desirable for analysis of cultural signals, since hesitation phenomena are important in language analyses.[12] For such purposes, the oral data would be used at normal speeds.

Finally, such findings as those concerning organizational effectiveness help clarify what may be the most lasting of the lessons taught by the evaluative experience described in this chapter--the lesson, namely, that the design is a frame of mind, in which phrase both nouns are important: "mind" effectively inhibiting any abdication to the numbers on the machines, and "frame" serving as a reminder that activities outside the scope of the design do not necessarily disappear because of that fact. For a more enduring statement of this lesson, the evaluators as behavioral scientists may find very useful the insights of the humanists (whom they too seldom consult): in his essay "The Figure a Poem Makes," Robert Frost observes: "Like giants we are always hurling experience ahead of us to pave the future with against the day when we may want to strike a line of purpose across it for somewhere." For people who gather and interpret data, it is the kind of insight that turns into advice.

FOOTNOTES

[1]W. Johnson, "A Program of Research," Psychological Monographs, 56 (1944), 1-15.

[2]Theodore Andersson and Mildred Boyer, Bilingual Schooling in the United States, Vol. 1 (Austin, Texas: Southwest Educational Development Laboratory, 1970), p. 67.

[3]Put differently, no specific interactive goals were developed prior to the start of the program, nor had measures been devised to determine their achievement. This is of course a very common situation in the evaluation arena.

[4]Some attempts have been made to determine the value of teachers in learning, obviously a difficult and highly charged question. See, for example, Eric Hunaschek,"Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data," The American Economic Review, 61 (May, 1971), 280-288.

[5]Philip Jackson, Life in Classrooms (New York: Holt, 1968); E. Webb, D.T. Campbell, R.D. Schwartz and L. Sechrest, Unobtrusive Measures (Chicago: Rand McNally, 1966).

[6]J. Grosswald, "Standardized Test Popularity" Measurement News (October, 1973), pp. 2-3.

[7]Robert Lado, Linguistics Across Cultures (Ann Arbor: University of Michigan Press, 1957); see also Lado's Language Teaching (New York: McGraw Hill, 1964).

[8]Andersson and Boyer, Bilingual Schooling in the United States, Vol. 1, p. 133.

[9]Carol H. Weiss, "The Politicization of Evaluation," Journal of Social Issues, 26 (Autumn, 1970), 64.

[10]Robert Rippey, Transactional Evaluation (Berkeley, California:, McCutchan, 1973).

[11]John J. Gumperz and Dell Hymes, Directions in Sociolinguistics: The Ethnography of Communications (New York: Holt, 1972), p. 24.

[12]Donald S. Boomer, "Hesitation and Grammatical Encoding," in The Psychosociology of Languages. ed. by Serge Moscovici (Chicago: Markham, 1972); Frieda Goldman-Eisler,"A Comparative Study of Two Hesitation Phenomena," Ibid.

SUGGESTED READINGS

Andersson, Theodore, and Boyer, Mildred. Bilingual Schooling in the United States. 2 vols. Austin, Texas: Southwest Educational Development Laboratory, 1970.

Boomer, Donald S. "Hesitation and Grammatical Encoding." The Psychosociology of Language. Edited by Serge Moscovici. Chicago: Markham, 1972.

Goldman-Eisler, Frieda. "A Comparative Study of Two Hesitation Phenomena." The Psychosociology of Language. Edited by Serge Moscovici. Chicago: Markham, 1972.

Grosswald, J. "Standardized Test Popularity." Measurment News. October, 1973, pp.2-3.

Gumperz, John J., and Hymes, Dell. Directions in Sociolinguistics: The Ethnography of Communication. New York: Holt, 1972.

Hunaschek, Eric. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." The American Economic Review, 61 (May, 1971) pp. 280-88.

Jackson, Philip. Life in Classrooms. New York: Holt, 1968.

Johnson, W. "A Program of Research." Psychological Monographs, 56 (1944), 1-15.

Lado, Robert. Linguistics Across Cultures. Ann Arbor: University of Michigan Press, 1957.

..... Language Teaching. New York: McGraw Hill, 1964.

Rippey, Robert. Transactional Evaluation. Berkeley: McCutchan, 1973.

Webb, E., Campbell, D.T., Schwartz, R.D., and Sechrest, L. Unobtrusive Measures. Chicago: Rand McNally, 1966.

Weiss, Carol H. "The Politicization of Evaluation." Journal of Social Issues, 26. (Autumn, 1970), 57-68.